

Notational Conventions

Zepu Zhang

August 29, 2010

(This note is not all understandable to you at the beginning of the course. Re-read it later.)

The textbook has several notational problems. Following the principles below, we will make some systematic changes to the notation used in the textbook.

All the problems stem from the fundamental distinction between random variable versus constants (fixed values, numbers), and between estimators and estimates.

It is conventional to use upper-case Roman letters, such as A, B, X, Y , for *variables*, and corresponding lower-case letters, such as a, b, x, y , for *specific values* these random variables happen to take (in any particular instance). It is also conventional to use Greek letters, such as α, β, θ , for *variables* or (unknown, to-be-estimated) constants.

Note the consistency in a formula. We may transform variable X to variable $Y = \exp(X)$; we may also convert number x to number $y = \exp(x)$.

When we talk about data (or dataset), there are two situations.

1. We have data Y_1, Y_2, \dots, Y_n . This means we are talking about a dataset with n entries. We treat the general situation that each $Y_i, i = 1, \dots, n$, may be any (reasonable) value. This is a *data variable*.

If we actually go to obtain one particular such dataset, the n values will be fixed numbers; but by using Y we do not care about this step.

In fact, if we actually obtain a specific value for each Y , there can be many possible values for each Y (and we'll happen to take one specific value). Therefore the Y 's are *random* variables.

2. We have data y_1, y_2, \dots, y_n . This means a dataset of n fixed numbers.

An “estimator” is a function of data. Here “data” are *random samples* from the data-generation model (for example $Y \sim N(\mu, \sigma^2)$). The data here are considered *random*, not yet fixed at the particular values observed in reality. Because the data are random, an estimator (which is a function of random data) is a random variable. An estimator is represented by a formula that involves random data (and possibly constants, such as coefficients).

The particular value of an estimator after plugging in a particular dataset is called the “estimate”. The estimate is simply a number; it's not a random variable.

In this class, the response is a random variable (being modeled), denoted by Y ; it also has particular values, denoted by y .

A problem of the textbook is to confuse the estimators and estimates of β . The same happens to σ^2 , as well as for the predictors and response.

We adopt a (hopefully) better notation scheme like the following.

1 The simple regression case

The predictor values are x_1, \dots, x_n . The predictor is a *variable* because it takes various values, but it is *not random*. Its values are fixed in a particular problem. When we refer to the “predictor variable” without meaning any particular value, we may use X . (I’m not sure I’ve found a really simple and consistent convention for this; but just remember it is *not random*.)

The response variable at x is Y . Corresponding to x_1 is Y_1 , to x_2 is Y_2 , etc. The actual set of observed values of the response variable is y_1, \dots, y_n .

In this course we assume the model correct describes the relation between the predictor and the response. The “model” here refers to (1) the formula of the model (the deterministic part); and (2) assumptions about how Y can deviate from the value predicted by the deterministic formula.

The fitted (i.e. estimated) value of Y at x , when we talk about the formula without plugging in a particular dataset, is \hat{Y} . Plugging a particular dataset y_1, \dots, y_n into the formula for \hat{Y} , we get \hat{y} . If the actual dataset is a bit different, then a different \hat{y} will result. But the general formula for \hat{Y} is the same.

The model parameters are $\beta_0, \beta_1, \sigma^2$. The estimators are $\hat{\beta}_0, \hat{\beta}_1, \hat{\sigma}^2$ or S^2 . The estimates are b_0, b_1, s^2 .

Example formulas:

$$\begin{aligned}\hat{\beta}_1 &= \frac{\sum(x_i - \bar{x})(Y_i - \bar{Y})}{\sum(x_i - \bar{x})^2}, & b_1 &= \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2}, & E(\hat{\beta}_1) &= \beta_1, & \text{var}(\hat{\beta}_1) &= \frac{\sigma^2}{\sum(x_i - \bar{x})^2} \\ \hat{\beta}_0 &= \bar{Y} - \hat{\beta}_1 \bar{x}, & b_0 &= \bar{y} - b_1 \bar{x} \\ \hat{Y}_i &= \hat{\beta}_0 + \hat{\beta}_1 x_i, & \hat{y}_i &= b_0 + b_1 x_i \\ S^2 &= \frac{\sum(Y_i - \hat{Y}_i)^2}{n - 2}, & s^2 &= \frac{\sum(y_i - \hat{y}_i)^2}{n - 2}, & s^2(\hat{\beta}_1) &= \frac{s^2}{\sum(x_i - \bar{x})^2}\end{aligned}$$

Note: only estimators have sampling distributions, expected values, variances, etc. These are meaningless for estimates, which are not random variables (they are just fixed numbers). The book is fundamentally confusing when it writes $b_1 = 3.1$ at one place, and $s^2\{b_1\} = 5$ at another. A fixed number does not have variance.

2 The multiple regression case

Each data point has a predictor *vector*, \vec{x} . For the i th observation, the predictor vector may be written \vec{x}_i .

The design matrix is \mathbf{X} . (Upper case, bold.)

The situation for the response is the same as in the simple regression case, because it is still a scalar. That is, corresponding to predictor vector \vec{x} is a scalar variable Y , or a single number y .

The model parameters are $\vec{\beta}$ (coefficient vector) and σ^2 . The estimators are $\hat{\vec{\beta}}$ and $\hat{\sigma}^2$ or S^2 . The estimates are \vec{b} and s^2 .

Example formulas:

$$\begin{aligned}\hat{\vec{\beta}} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\vec{Y}, & \vec{b} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\vec{y}, & E(\hat{\vec{\beta}}) &= \vec{\beta}, & \text{cov}(\hat{\vec{\beta}}) &= \sigma^2(\mathbf{X}'\mathbf{X})^{-1} \\ \hat{\vec{Y}} &= \mathbf{X}\hat{\vec{\beta}}, & \hat{\vec{y}} &= \mathbf{X}\vec{b} \\ S^2 &= (n-2)^{-1}(\vec{Y} - \hat{\vec{Y}})'(\vec{Y} - \hat{\vec{Y}}), & s^2 &= (n-2)^{-1}(\vec{y} - \hat{\vec{y}})'(\vec{y} - \hat{\vec{y}}), & s^2(\hat{\vec{\beta}}) &= s^2(\mathbf{X}'\mathbf{X})^{-1}\end{aligned}$$

I believe the arrow symbol for vector has its origin in hand-writing. In print we can use boldness (a similar situation is italic in print and underlining in hand-writing for emphasis). The above may be re-written as follows.

$$\begin{aligned}\hat{\boldsymbol{\beta}} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}, & \mathbf{b} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}, & E(\hat{\boldsymbol{\beta}}) &= \boldsymbol{\beta}, & \text{cov}(\hat{\boldsymbol{\beta}}) &= \sigma^2(\mathbf{X}'\mathbf{X})^{-1} \\ \hat{\mathbf{Y}} &= \mathbf{X}\hat{\boldsymbol{\beta}}, & \hat{\mathbf{y}} &= \mathbf{X}\mathbf{b} \\ S^2 &= (n-2)^{-1}(\mathbf{Y} - \hat{\mathbf{Y}})'(\mathbf{Y} - \hat{\mathbf{Y}}), & s^2 &= (n-2)^{-1}(\mathbf{y} - \hat{\mathbf{y}})'(\mathbf{y} - \hat{\mathbf{y}}), & s^2(\hat{\boldsymbol{\beta}}) &= s^2(\mathbf{X}'\mathbf{X})^{-1}\end{aligned}$$

Again, it makes no sense to write $s^2(\mathbf{b})$, because \mathbf{b} is not random.

Understand \mathbf{Y} : it's a random vector. Vector, hence bold; random, hence upper case. It means the vector with elements Y_1, \dots, Y_n . Distinguish \mathbf{Y} from the design matrix \mathbf{X} , which is a matrix constant.

Another notation is the critical values $t(1 - \alpha/2; \nu)$ and $F(1 - \alpha; \nu_1, \nu_2)$. They are more commonly written as $t(\alpha/2; \nu)$ and $F(\alpha; \nu_1, \nu_2)$, or $t_{\alpha/2; \nu}$ and $F_{\alpha; \nu_1, \nu_2}$. It is more conventional to use the *right tail* (i.e. α or $\alpha/2$) to identify the critical value.